



A Review of Metaphase Image Selection Techniques for Automatic Karyotype Generation

Tanvi Arora¹ and Renu Dhir²

¹Department of Computer Science & Engineering, Baddi University of Emerging Sciences & Technology, Baddi, Solan, Himachal Pradesh, India.

²Department of Computer Science & Engineering, National Institute of Technology, Jalandhar, Punjab, India.

ARTICLE INFO

Article history:

Received: 6 January 2015;

Received in revised form:

15 March 2015;

Accepted: 9 April 2015;

Keywords

Chromosomes,
Karyotype,
Metaphase Images,
Genetic disorders.

ABSTRACT

The karyotype is analyzed to detect the genetic abnormalities. It is generated by arranging the chromosomes after extracting them from the metaphase images. The chromosomes are non-rigid bodies that contain the genetic information of an individual. The metaphase spread contains the chromosomes but these chromosomes are not distinct bodies, they can either be individual chromosomes or they may be touching one another, they may be bent or even may be overlapping and thus forming a cluster of chromosomes. The extraction of these touching and overlapping chromosomes is a very tedious process. The segmentation of these may not give us correct and accurate results. Therefore before taking up a metaphase image for analysis it must be analyzed for the orientation of the chromosomes it contains. The various reported methods for metaphase image selection for automatic karyotype generation are compared in this paper, the images of metaphase spread are having hypo or hyper fluorescent regions and there is variability of contrast between the background and the chromosomes. After analysis it has been concluded that each metaphase image selection method has its advantages and disadvantages. The MetaSel software outperforms all the methods and is having the overall best performance.

© 2015 Elixir All rights reserved.

Introduction

The cytogenetic analysis is a two-step process requiring metaphase image selection and then karyotype analysis[5]. The chromosomes carry the genetic information of an individual, various genetic disorders are discovered by studying the chromosomes. Each healthy human cell has 23 pairs of chromosomes, comprising of 22 pairs of autosomes and a pair of sex chromosomes[5]. Karyotypes are studied to find out the genetic abnormalities, there may be some missing chromosome or an extra chromosome or a deletion of a part of a chromosome or duplication.

The karyotype of human chromosomes are prepared by segmenting the chromosomes present in metaphase Images. The metaphase images are taken using a microscope. Before taking an image of a metaphase spread the chromosomes are stained with a fluorescent dye. The blood sample of an individual is taken, approximately 10 glass slides are pre-pared and approximately 20 metaphase spreads are present on a single glass slide. The chromosomes appear as distinguishable individual bodies only at the end of the cell division. They appear as long objects resembling the strings.

The first step in karyotype generation is the selection of the metaphase image from the set of images that are generated. Approximately 200 metaphase spread images are available. For manual analysis by an experienced cytogeneticist 20 best images are selected. The cytogeneticist then visually inspects each of the 20 selected metaphase images. The images are inspected to see the number of chromosomes present in the metaspread. Once this visual inspection is over then out of 20, approximately 5 sharp images are chosen for analysis purpose. All the tasks are time consuming and they require the experienced person to visually inspect and select the best metaphase image.

The metaphase image that are selected are used for segmentation of the chromosomes and then further each segmented chromosome is classified into one of the 22 classes or X or Y sex chromosomes. Many automatic tools for segmenting and classifying the chromosomes have been developed. But the big challenge for all these tools is that the chromosomes are non-rigid objects and they do not always occur straightened, but they are either present in different orientations, or they may be bent or they may be touching one another or they are overlapping thus forming clusters.

In addition to this the metaphase images may also contain some artifacts that are not chromosomes. So all these things make the task of automatic segmentation and classification of human chromosomes into a karyotype a tedious one[3]. In order to get an accurate karyotype the metaphase image that is selected for segmentation and classification of chromosomes should have more of individual chromosomes, that have clear band patterns and have straight orientations and also they are not touching or overlapping.

To automate the process of karyotyping hundreds of metaphase images needs to be physically examined by an experienced cytogeneticist, which may require a lot of time and effort[4]. And the mood of the cytogeneticist may also effect the selection of the metaphase images. So a cytogeneticist may select 2 best images and leave the remaining images that may contain vital information, so this method of manual selection may not result in selection of the best metaphase images. In this study the various automatic techniques for automatic metaphase image selection are studied and compared.

The paper is organized as follows section 2 contains the various methods for automatic metaphase image selection,

section 3 contains the comparative analysis section 4 contains conclusion.

Methods for Metaphase Image selection

R Huber's Method

Finding a metaphase is a tedious and time consuming job. But the job of metaphase finding or choosing is less complex as compared to chromosome analysis and karyotyping and can be automated to great extent. The Metafer 2 system is hardware and software based system [8]. The detection of a metaphase is done using a three phase algorithm. In the first phase the line values of band pass filtered objects count (FOC) are analyzed, if it has consecutive high value it indicates a candidate metaphase. In the second phase the horizontal distribution of the FOC is determined for these lines from the binary image. So here a sequence of high column values is taken as candidate metaphase thereby reducing the area of interest to a rectangle. In the third pass the binary image located in the rectangle is fully analyzed by a rapid contour following algorithm. A total of nine features are computed and a multi variant statistical classifier is used for the final classification as a metaphase or a non-metaphase. the classification is based upon the binary images before extracting the features thresholding is applied to remove the background and thus reducing the data to be analyzed.

The nine features that are extracted are:

- 1)The sum of the band pass filtered FOC values within the rectangle
- 2)Number of objects within the rectangle which are larger than the given threshold value.
- 3)The mean area of these objects
- 4)The mean contour length
- 5)Mean quotient of area and contour length
- 6)Square of total contour length divided by total area
- 7)The sum of the product of the object area and the distance of the object from center of all objects
- 8)Sum of the distances of the objects from the center
- 9)Sum of the center distances divided by the sum of the squared center distances of the objects

Then the multivariate classifier is used to classify the metaphase and non-metaphase

Victor Gajendran's Method

In this work , system is proposed that automatically counts the chromosomes in digital images [2]. This system has two phases namely preprocessing and counting phase. In the preprocessing step hysteresis thresholding segments the chromosome objects from the background. Further median filtering method is used to remove salt and pepper noise and also filling the holes of chromosomes and to smoothen the chromosome contours so that when thinning is applied then extra branches are not created. After this thinning operation is performed to obtain the single pixel width skeletons of chromosomes or their clusters. After this the average width of all the skeletons is calculated. It has been observed that all the chromosomes have consistent width. So all those skeletons who are less than the average width of all the skeletons are treated as noise and are not considered. Based upon the same lines the slight connections are also removed.

The counting phase takes the noise free and skeletonized metaphase images as input and it labels all the skeletons present in the input metaphase image. Then for each component the end points and cross overs are identified. The raster scanning method is used for finding the first end point and it is traced until next end point is reached and if a crossover is reached then the path that comes first appears is taken. While tracking the skeleton all other pixels are deleted except for the crossover pixels. The count of chromosomes is incremented by one for each

component and all the above steps are repeated till all the end points are traced.

Wang X's Method

This method classifies the metaspread images into two categories of analyzable metaphase cells and unanalyzable metaphase cells [1]. It is a five step process. In the first step they take up the digital image and the image quality is enhanced using median filtering. In the second step the thresholding is applied to remove the high grey values. Third step is region labeling to find connected components and delete the isolated pixels. The fourth step computes the five image features from the labeled components. The five features to be computed are:

1. The count of the labeled components (N_m) is computed for each metaspread image.
2. The pixels present in each labeled component is counted ($S_i \times N_m$)
3. The circularity of each labeled component is calculated as $C_i = \frac{N_c}{N_m}$, where N_c is the number of pixels that are located inside the region contour and circle and N_m is count of pixels inside the labeled component.
4. Each labeled components average grey value is computed as ($I_{ave} = \frac{1}{n} \sum_1^n I_i$)
5. The length between the center of the cell (x_c, y_c) and each labeled region (x_k, y_k) is calculated.

$$L_k = \sqrt{(x_c - x_k)^2 + (y_c - y_k)^2}$$

Then in the fifth step the computed features are passed to two machine learning based classifiers namely decision trees and artificial neural networks and the results are further optimized.

Yan Wenzhong's Method

In this work an attempt has been made to extract the count of chromosomes from a metaphase image that has overlapping chromosomes. For selecting the overlapping chromosomes images firstly they have used histogram equalization to improve the contrast of the image [3]. After histogram equalization the image was transformed into binary image by using thresholding. Then the binary image was eroded in order to delete the small sized light colored objects.

Let A_k be an image that is being eroded kth time

$$A_k = A \otimes kB$$

Where B represents structure element, \otimes Denote erosion element. Y_k denotes a subset of ultimate connected components in one element of A_k if $l > k$. A_k will disappear from A_l

$$U_k = (A_{k+1} \oplus \{B\}); A_k$$

Then $Y_k = A_k - U_k$, Where \oplus denotes dilation operator

In case there are many objects in the image then ultimate connected components of image are denoted as

$$Y = \bigcup_{k=1, m} Y_k$$

Where m denotes the number of times erosion operation is applied. After erosion 8 connectivity labeling algorithm was used to sign exclusive labels to every object in the image.

After this the overlapping chromosomes were counted by the following method:

1. Initially it is assumed that there is only one chromosomes in the cluster.

Max=1

2. Then a structure element B is taken with Euclid disc whose radius is 3

3. The image is eroded with B

4. Then the eroded image is labeled using 8 connected component labeling algorithm and m is used to represent the number of chromosomes in the eroded image.
5. If the $num \neq 0$, num will be compared with max .
6. If $num > max$ then $max = num$
7. But if num is not greater than max then we have to go to step 3 for the next loop
8. Otherwise when $num = 0$ the value of max is just number of chromosomes counted in the overlapping chromosome image.

Uttamatatin's MetaSel Method

They have taken the metaphase images and prior to analysis they have defined the objects of the metaphase spread into four different classes. The class 1 is assigned to objects that are straight and are separate distinguishable entities, class 2 is assigned to objects that are either bent or skewed entities, class 3 is assigned to objects that are clusters of touching or overlapping entities, whereas class 4 is assigned to left overs of the cell or artifacts[5].

There work aims at automatically choosing a high quality metaphase spread image so as to make the task of automatic karyotyping simpler and more accurate. They have preprocessed the metaphase images using Otsu's automatic thresholding [9] to segment the chromosome images from the background. The classification of the chromosomes has been done using width, height and its ratio so as to categorize the objects into four classes as defined above.

$$Area_{ratio} = \frac{A_o}{A_r}$$

Where A_r is count of pixels in the enclosing rectangle ($W_{rect} \times H_{rect}$) of segmented object and A_o is the number of pixels of the segmented objects. W_{rect} and H_{rect} are the width and height of the enclosing rectangle. Area ratio can be used to check if the chromosome is straight. The empirical probability density function was calculated using the kernel density method. Then a Gaussian based model was used to find the minimum threshold value for the ratio of area so as to classify the chromosome into class 1. But this may also contain some cell residue or artifacts. In order to remove them it was observed that the width of the class 1 chromosomes is consistent, so the objects classified as class 1 can be considered as cell residues or artifacts if their width is not consistent and they can be assigned class 4. So if the width of the object is 1.5 times then the average width of the class 1 objects, then they are discarded.

Let O_w represent the set of objects with the width less than 1.5 times the total average width. The width of each object W in set O_w can be defined as

$$W_i = \frac{\text{Total number of pixels in chromosome}}{H_{rect}}$$

/

Then average width is

$$W_{avg} = \frac{\sum_{i \in O_w} W_i}{|O_w|}$$

To quantify the deviation from the average width we define the rectangle width ratio as

$$W_{rect_{ratio}} = \frac{W_{rect}}{W_{avg}}$$

So if $W_{rect_{ratio}}$ for an object is less than the threshold value then it can be classified into class 2, class 3 or class 4 categories

Now in order to distinguish between these classes. They took the height of the segmented object as parameter.

$$H_i = \frac{A_o}{W_{avg}}$$

The ratio between H_i and H_{rect} , height ratio ($H_{i_{ratio}}$) is compared by

$$H_{i_{ratio}} = \frac{H_i}{H_{rect}}$$

Using empirical density function and Gaussian model it was found that the objects having $H_{i_{ratio}}$ less than a threshold value were considered as class 4 objects.

Now in order to distinguish between class 2 and class 3 objects one more parameter was computed and is termed as maximum width ratio ($W_{max_{ratio}}$).

$$W_{max_{ratio}} = \frac{W_{max}}{W_{avg}}$$

The empirical probability density function and Gaussian model were used to find the threshold value separating the class 2 and class 3 objects. When $W_{max_{ratio}}$ will be greater than threshold the object will be classified as class 3 object and the one with lesser will be classified as class 2 object.

The simple rule based Gaussian classification techniques can rank the metaspread images depending upon the number of objects of each class. The metaspread images having highest number of class 1 objects are the best candidates to be chosen for further automatic segmentation and classification of chromosomes for karyotype generation.

Ravi Uttamatatin's Band Classification Method

In this work they have classified the chromosome metaphase images into low and high band resolution considering the shape of chromosomes [7]. In the low resolution band chromosomes are small in size and are well spread and there is no touching or overlapping, so it is suitable for counting the number of chromosomes. In the case of high band resolution the chromosomes are long, they may be bent or overlapping so these band chromosomes are used for detecting structural abnormalities. The metaphases with low band resolution are used for counting the number of chromosomes whereas the high band resolution images are considered for structural abnormality checking.

In order to classify the metaphase images based upon resolution of banding the metaphase images are preprocessed based upon grey level adjustment and Otsu's thresholding to separate the fore ground and back ground. After segmenting the foreground and back ground the segmented objects are rotated so that they are vertical. After this preprocessing steps the parameters of the individual chromosomes are calculated such as area ratio ($Area_{ratio}$), Average width (W_{avg}), Width of each object ($Width(W_i)$), ratio of width and average width ($W_{rect_{ratio}}$), height (H_i), height ratio ($H_{i_{ratio}}$), maximum width ratio ($W_{max_{ratio}}$), Length (L).

Then based upon the following algorithm the chromosomes are classified as low band resolution and high band resolution.

1. Image parameters such as area ratio ($Area_{ratio}$), Average width (W_{avg}), Width of each object ($Width(W_i)$), ratio of width and average width ($W_{rect_{ratio}}$), height (H_i), height ratio ($H_{i_{ratio}}$), maximum width ratio ($W_{max_{ratio}}$), Length (L) are calculated

2. If the area ratio > threshold value the objects are classified as other objects else they are classified as straight chromosomes.

Table 1. Comparative Analysis of Metaphase Image Selection Techniques

S.no	Feature	R Huber Method[8]	Victor Gajendran Method[2]	X Wang Method [1]	Yan Wenzhang Method[3]	Ravi Uttamatinin MetaSel Method[5]	Ravi Uttamatinin Band Classification Method [7]
1	Number of classes	2	0	2	0	4	2
2	control over quality	No	No	No	No	Yes	No
3	Speed	Slow	Slow	Slow	Fast	Fast	Fast
4	Does it Rank Images	No	No	No	No	Yes	No
5	Features considered	Geometric	Geometric	Geometric	Geometric	Geometric	Shape & Band Features
6	Error rate	Not specified	6%	Not Given	Not Given	10%	15%
7	Counts number of chromosomes	No	Yes	No	Yes	Yes	No
8	counts number of overlapping chromosomes	No	Yes	No	No	Yes	No
9	Considers individual chromosome objects	No	Yes	No	Yes	Yes	Yes
10	metaspread images can be used for any method.	No	No	Yes	No	Yes	Yes
11	Workable in case of touching and overlapping chromosomes.	No	Yes	Yes	No	Yes	Yes
12	Workable when the chromosomes are thin and long.	No	Yes	Yes	No	Yes	Yes

For straight chromosomes W_{avg} is calculated, and then compared with min W_{rect_ratio} if it is less than threshold then the objects are classified as small artifacts and if that is more than the max threshold then they are classified as large artifacts else they are classified as individual chromosomes.

3. For other objects H_{i_ratio} is calculated if it is less than a threshold value then they are classified as small artifacts else the W_{max_ratio} is calculated if it is greater than the threshold then the objects are classified as large artifacts, otherwise they are classified as individual chromosomes.

4. For the individual chromosomes the length is calculated

5. If length is greater than a threshold value then the image is classified as high band resolution image else it is classified as low band resolution image.

Comparative Analysis

The task of automatic karyotyping of human chromosomes is very challenging. The chromosomes are non-rigid bodies which are scanned using a microscope and are visible as distinct bodies only towards the end of cell division. The metaspread thus images contains the visible chromosomes. But these chromosomes may be bent, touching or overlapping. Due to this studying the characteristic features of an individual chromosome is a tedious process. After 40 years of long research the task of automatic chromosome segmentation and classification is still an open issue.

After reviewing the literature and work done by various researchers in the field of automatic karyotyping, it has been concluded that owing to the non-rigid nature of chromosomes, the segmentation of touching and overlapping chromosomes cannot guarantee 100% accurate feature extraction by automated means.

So it is suggested to take up those metaspread images only for analysis or study purpose for automatic karyotyping that has least amount of bent, touching or overlapping chromosomes, so that the chance of inaccuracies are minimized.

Most of the works for automatic segmentation and classification of human chromosomes rely on manual selection of metaspread images. The manual selection is very time consuming process. The efficiency of the process depends upon the human behavior and the time constraints. So the search results may be biased or not up to the mark or it may be that whole of the search space is not explored.

So there is a strong need to automate the process of metaspread image selection prior to automated segmentation and classification of the chromosomes. Few researchers have worked in this area and have come forward with various methods by which they can classify the metaspread images into analyzable and non-analyzable categories, some have tried to rank the metaspread images in order of most analyzable to last analyzable depending upon the number of individual chromosomes, bent, touching or overlapping chromosomes present in the metaspread.

To date very less focus has been given to automate the process of automatic metaspread selection. The automatic selection of metaspread images will benefit in two ways; firstly it will reduce the time of cytogeneticist to select the metaspread image manually and other is the time saved while processing and analyzing the metaphase image for segmentation and classification purpose. R Huber's method [8] generates two classes and classifies the image as either as metaspread image or a non metaspread image, it has no control over quality of metaspread images, it is a slow process, no ranking is done, it considers only geometric features. Victor Gajendran's method [2], helps to count number of chromosomes, no control on quality, slow process, no ranking is done, no classification is done it just counts number of overlapping chromosomes by considering only geometric features. X Wang's method [1] just finds if the metaspread image is analyzable or not, does not consider quality, do not consider individual chromosome objects, no ranking, considers only geometric features. Yan Wenzhang's method [3] counts chromosomes, not workable in

case of touching and overlapping chromosomes, not workable when the chromosomes are thin and long, does not consider quality, does not rank or classify, considers only geometric features, is fast and simple. Ravi Uttamatin's MetaSel method [5] More accurate results as they classify and rank metaspread images, fast, task of segmentation of touching, bent or overlapping chromosomes will be reduced, considers quality, considers only geometric features. Ravi Uttamatin's Band classification method [7] classify into two groups as low band resolution and high band resolution, based upon shape features the band resolution information is obtained, low band resolution is used for counting the chromosomes whereas high band resolution is used for finding structural anomalies. Every proposed method has its own set of features. In the below Table 1, all the methods are compared against a set of features. The comparative analysis suggest that the best method so far is the Ravi Uttamatin's MetaSel method, as it ranks the metaspread images based upon the objects present, thereby making the task of karyotype generation very easy and fast. Thus saving a considerable time in manual selection and ranking of the metaspread images before using them for the segmentation purpose. The further work carried by Ravi Uttamatin based upon Band classification is also good, but it considers only single parameter that is the length of the chromosome to classify the metaphase images into low band and high band. Other methods are not so efficient as they just either counts the chromosomes or they just suggest whether the metaspread images are analyzable or not, without considering the quality of the images.

Conclusions

Based upon the comparative analysis it can be concluded that the Ravi Uttamatin's MetaSel[5] method is the best possible approach so far as it can be used to rank the metaspread images based upon the number of fully disentangled and non-overlapping chromosomes. This method further helps in reducing the complexity of automatic segmentation and classification and make the task simpler and efficient. The second method of classification based on band classification is also a suitable candidate, but it just takes one parameter i.e length feature to classify so that may not always yields accurate results.

Both the methods consider the quality of the metaphase images and tries to remove the small and large artifacts that are captured during microscopic images of the metaspreads. Thus automation of the metaspread image selection process can further enhance the efficiency of the automatic karyotype generation process and the reliance on the experienced cytogeneticist is also removed. There by creating a fully automatic system for automatic karyotype generation.

References

- Wang X, Li S, Liu H, Wood M, Chen W, Zheng B: Automated identification of analyzable metaphase chromosomes depicted on microscopic digital images. *Journal of biomedical informatics* 2008, 41(2):264-271.
- Gajendran V, Rodríguez J: Chromosome counting via digital image analysis. In *Proceedings of the 2004 International Conference on Image Processing (ICIP 2004)*. Edited by Institute of Electrical and Electronics Engineers (IEEE). Singapore; 2004:2929-2932. 24-27 October 2004
- Wenzhong Y: A Counting Algorithm for Overlapped Chromosomes. In *3rd International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2009)*. Edited by Institute of Electrical and Electronics Engineers (IEEE). Beijing, China; 2009:1-3. 11-13 June 2009
- Kovács G, Kajtár B, Méhes G, Fazekas A: Fast detection of chromosome metaphases in digitalized microscopic slides. In *International Conference on Image Analysis and Signal Processing (IASP)*. Edited by Institute of Electrical and Electronics Engineers (IEEE). Linhai, China; 2009:444-448. 11-12 April 2009
- Uttamatin R, Yuvapoositanon P, Intarapanich A, Kaewkamnerd S, Phuksaritanon R, Assawamakin A, Tongshima S, MetaSel: a metaphase selection tool using a Gaussian-based classification technique. *BMC Bioinformatics*. 2013;14 Suppl 16:S13. doi: 10.1186/1471-2105-14-S16-S13. Epub 2013 Oct 22.
- Uttamatin, R.; Yuvapoositanon, P.; Intarapanich, A.; Kaewkamnerd, S.; Tongshima, S., "Chromosome classification for metaphase selection," *Communications and Information Technologies (ISCIT)*, 2013 13th International Symposium on ,vol., no., pp.464,468, 4-6 Sept. 2013,doi: 10.1109/ISCIT.2013.6645903
- Uttamatin, R.; Intarapanich, A.; Kaewkamnerd, S.; Yuvapoositanon, P.; Tongshima, S., "Band classification based on chromosome shapes," *Biomedical Engineering International Conference (BMEiCON)*, 2013 6th , vol., no., pp.1,5, 23-25 Oct. 2013, doi: 10.1109/BMEiCon.2013.6687672
- R. Huber and U. Kulka and T. Lrch and H. Braselmann and M. Bauchinger, Automated metaphase finding: an assessment of the efficiency of the METAFER2 system in routine mutagenicity assay, *PubMed, Mutation Research*, 1995, vol 334 97- 102
- Liu D, Yu J: Otsu method and K-means. In *Proceedings of the Ninth International Conference on Hybrid Intelligent Systems (HIS 2009)*. Shenyang, China; Institute of Electrical and Electronics Engineers (IEEE) 2009:344-349, 12-14 August 2009.
- Truong K, Gibaud A, Zalcman G, Guilly M. Quantitative fish determination of chromosome 3 arm imbalances in lung tumors by automated image cytometry. *Med Sci Monit* 2004;10:426-32.
- Wang X, Zheng B, Wood M, Li S, Chen W, Liu H. Development and evaluation of automated systems for detection and classification of banded chromosomes: current status and future perspectives. *J Phys D Appl Phys* 2005;38:2536-42.
- Popescu M, Gader P, Keller J, Klein C. Automatic karyotyping of metaphase cells with overlapping chromosomes. *Comput Biol Med* 1999;29:61-82.
- Cosio F, Vega L, Becerra A, Melendez C. Automatic identification of metaphase spreads and nuclei using neural networks. *Med Biol Eng Comput* 2001;39:391-6.
- Castleman KR. The PSI automatic metaphase finder. *J Radiat Res* 1992;33:124-8.
- Piper J, Granum E. On fully automatic feature measurement for banded chromosome classification. *Cytometry* 1989;10:242-55.
- Errington P, Graham J. Application of artificial neural networks to chromosome classification. *Cytometry* 1993;14:627-39